944-001.051

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : H04N 17/00 — A1

(11) International Publication Number: **WO 00/48407**

(43) International Publication Date: 17 August 2000 (17.08.00)

(21) International Application Number: PCT/GB00/00171

(22) International Filing Date: 24 January 2000 (24.01.00)

(30) Priority Data:

9903107.2	11 February 1999 (11.02.99)	GB
9903293.0	12 February 1999 (12.02.99)	GB
99304824.8	18 June 1999 (18.06.99)	EP

(71) Applicant *(for all designated States except US)*: BRITISH TELECOMMUNICATIONS PUBLIC LIMITED COMPANY [GB/GB]; 81 Newgate Street, London EC1A 7AJ (GB).

(72) Inventors; and

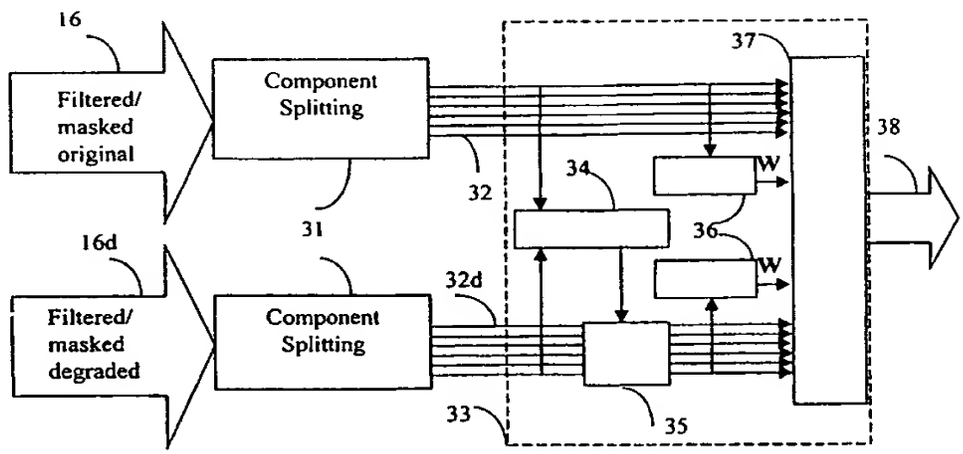
(75) Inventors/Applicants *(for US only)*: HOLLIER, Michael, Peter [GB/GB]; 4 Farlingayes, Woodbridge, Suffolk IP12 1HF (GB). BOURRET, Alexandre [FR/GB]; 30 Finchley Road, Ipswich, Suffolk IP4 2HU (GB).

(74) Agent: LIDBETTER, Timothy, Guy, Edwin; BT Group Legal Services, Intellectual Property Department, 8th floor, Holborn Centre, London EC1N 2TE (GB).

(81) Designated States: CA, JP, US, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(54) Title: ANALYSIS OF VIDEO SIGNAL QUALITY



(57) Abstract

Two video signals, typically an original signal (16) and a degraded version (16d) of the same signal, are analysed firstly to identify the perceptually relevant boundaries of the elements forming the video images depicted therein (31). These boundaries are then compared (33) to determine the extent to which the properties of the boundaries defined in one image (16) are preserved in the other (16d), to generate an output (38) indicative of the perceptual difference between the first and second signals. The boundaries may be defined by edges, colour, luminance or texture contrasts, disparities between frames in a moving or steroscopic image, or other means. The presence, absence, difference in clarity or difference in means of definition of the boundaries is indicative of the perceptual importance of the differences between the signals, and therefore of the extent to which any degradation of the signal (16d) will be perceived by the human viewer of the resulting degraded image. The results may also be weighted (36) according to the perceptual importance of the image depicted – for example the features which identify a human face, and in particular those responsible for visual speech cues.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

ANALYSIS OF VIDEO SIGNAL QUALITY

This invention relates to the analysis of the quality of video signals. It has a number of applications in monitoring the performance of video transmission equipment, either during development, under construction, or in service.

As communications systems have increased in complexity it has become increasingly difficult to measure their performance objectively. Modern communications links frequently use data compression techniques to reduce the bandwidth required for transmission. When signals are compressed for more efficient transmission, conventional engineering metrics, such as signal-to-noise ratio or bit error rate, are unreliable indicators of the performance experienced by the human being who ultimately receives the signal. For example, two systems having similar bit-error rates may have markedly different effects on the quality of the data (sound or picture) presented to the end user, depending on which digital bits are lost. Other non-linear processes such as echo cancellation are also becoming increasingly common. The complexity of modern communications systems makes them unsuitable for analysis using conventional signal processing techniques. End-to-end assessment of network quality must be based on what the customer has, or would have, heard or seen.

The main benchmarks of viewer opinion are the subjective tests carried out to International Telecommunications Union standards P.800, *"Methods for subjective determination of transmission quality", 1996* and P.911 *"Subjective audiovisual quality assessment methods for multimedia applications", 1998.* These measure perceived quality in controlled subjective experiments, in which several human subjects listen to each signal under test. This is impractical for use in the continuous monitoring of a network, and also compromises the privacy of the parties to the calls being monitored. To overcome these problems, auditory perceptual models such as those of the present applicant's International Patent Specifications WO 94/00922, WO95/01011, WO95/15035, WO97/05730, WO97/32428, WO98/53589 and WO98/53590 are being developed for measuring telephone network quality. These are objective performance metrics, but are designed to relate directly to perceived signal quality, by producing quality scorings similar to those which would have been reported by human subjects.

The prior art systems referred to above measure the quality of sound (audio) signals. The present invention is concerned with the application of similar principles to video signals. The basic principle, of emulating the human perceptual system (in this case the eye/brain system instead of the ear/brain system) is still used, but video signals and the human visual perceptual system are both much more complex, and raise new problems.

As with hearing, the human visual perception system has physiological properties that make some features present in visual stimuli very difficult or impossible to perceive. Compression processes, such as those established by JPEG (Joint Pictures Expert Group) and MPEG (Motion Pictures Expert Group) rely on these properties to reduce the amount of information to be transmitted in video signals (moving or still). Two compression schemes may result in similar losses of information, but the perceived quality of a compressed version of a given image may be very different according to which scheme was used. The quality of the resulting images cannot therefore be evaluated by simple comparison of the original and final signals. The properties of human vision have to be included in the assessment of perceived quality.

It is problematic to try and locate information from an image by mathematical processing of pixel values. The pixel intensity level becomes meaningful only when processed by the human subject's visual knowledge of objects and shapes. In this invention, mathematical solutions are used to extract information resembling that used by the eye-brain system as closely as possible.

A number of different approaches to visual modelling have been reported. These are specialised to particular applications, or to particular types of video distortion. For example, the MPEG compression system seeks to code the differences between successive frames. At periods of overload, when there are many differences between successive frames, this process reduces the pixel resolution, causing blocks of uniform colour and luminance to be produced. Karunasekera, A. S., and Kingsbury, N. G., in *"A distortion measure for blocking artefacts in images based on human visual sensitivity", IEEE Transactions on Image Processing, Vol. 4, No. 6, pages 713-724, June 1995,* propose a model which is especially designed to detect "blockiness" of this kind. However, such blockiness does not always signify an error, as the effect may have been introduced deliberately by the producer of the image, either for visual

effect or to obliterate detail, such as the facial features of a person whose identity it is desired to conceal.

If the requirements of a wide range of applications, from high definition television to video conferencing and virtual reality, are to be met, a more complex architecture has to be used.

Some existing visual models have an elementary emulation of perceptual characteristics, referred to herein as a "perceptual stage". Examples are found in the Karunasekera reference already discussed, and Lukas, X. J., and Budrikis, Z. L., *"Picture Quality Prediction Based on a Visual Model", IEEE Transactions on Communications, vol. com-30, No. 7, pp. 1679-1692 July 1982,* in which a simple perceptual stage is designed around the basic principle that large errors will dominate subjectivity. Other approaches have also been considered, such as a model of the temporal aggregation of errors described by Tan, K. T., Ghanbari, M. and Pearson, D. E., *"A video distortion meter", Informationstechnische Gesellschaft, Picture Coding Symposium, Berlin, September 1997.* However, none of these approaches addresses the relative importance of all errors present in the image.

For the purposes of the present specification, the "colour" of a pixel is defined as the proportions of the primary colours (red, green and blue) in the pixel. The "luminance" is the total intensity of the three primary colours. In particular, different shades on a grey scale are caused by variations in luminance.

According to a first aspect of the present invention, there is provided a method of measuring the differences between a first video signal and a second video signal, comprising the steps of:

analysing the information content of each video signal to identify the perceptually relevant boundaries of the video images depicted therein;

comparing the boundaries so defined in the first signal with those in the second signal; the comparison including determination of the extent to which the properties of the boundaries defined in the original image are preserved, and

generating an output indicative of the perceptual difference between the first and second signals.

According to a second aspect of the present invention, there is provided apparatus for measuring the differences between a first video signal and a second video signal, comprising:

analysis means for the information content of each video signal, arranged to identify the perceptually relevant boundaries of the video images depicted therein;

comparison means for comparing the boundaries so defined in the first signal with those in the second signal; the comparison including determination of the extent to which the properties of the boundaries defined in the original image are preserved,

and means for generating an output indicative of the perceptual difference between the first and second signals.

The boundaries between the main elements of an image may be identified by any measurable property used by the human perceptual system to distinguish between such elements. These may include, but are not limited to, colour, luminance, so-called "hard" edges (a narrow line of contrasting colour or luminance defining an outline or other boundary, such a line being identifiable in image analysis as a region of high spatial frequency), and others which will be discussed later.

The properties of the boundaries on which the comparison is based include the characteristics by which such boundaries are defined. In particular, if a boundary is defined by a given characteristic, and that characteristic is lost in the degraded image, the degree of perceived degradation of the image element is dependant on how perceptually significant the original boundary was. If the element defined by the boundary can nevertheless be identified in the degraded image by means of a boundary defined by another characteristic, the comparison also takes account of how perceptually significant such a replacement boundary is, and how closely its position corresponds with the original, lost, boundary.

The basis for the invention is that elements present in the image are not of equal importance. An error will be more perceptible if it disrupts the shape of one of the essential features of the image. For example, a distortion present on an edge in the middle of a textured region will be less perceptible than the same error on an independent edge. This is because an edge forming part of a texture carries less information than an independent edge, as described by Ran, X., and Favardin, N., *"A Perceptually Motivated Three-Component Image Model - Part II: Application to Image Compression", IEEE Transactions on Image Processing, Vol. 4, No. 4, pp. 713-724, April 1995*. If, however, a textured area defines a boundary, an error that changes the properties of the texture throughout the textured area can be as important as an error on an independent edge, if the error causes the textured characteristics of the

area to be lost. The present invention examines the perceptual relevance of each boundary, and the extent to which this relevance is preserved.

The process identifies the elements of greatest perceptual relevance, that is the boundaries between the principal elements of the image. Small variations in a property within the regions defined by the boundaries are of less relevance than errors that cause the boundary to change its shape.

Moreover, the process allows comparison of this information independently of how the principal elements of the images are identified. The human perceptual system can distinguish different regions of an image in many different ways. For example, the absence of a "hard edge" will create a greater perceptual degradation if the regions separated by that edge are of similar colour than it will if they are of contrasting colours, since the colour contrast will still allow a boundary to be perceived. The more abrupt the change, the greater the perceptual significance of the boundary.

By analysing the boundaries defined in the image, a number of further developments become possible.

The boundaries can be used as a frame of reference, by identifying the principal elements in each image and the differences in their relative positions. By using differences in relative position, as opposed to absolute position, perceptually unimportant differences in the images can be disregarded, as they do not affect the quality of the resulting image as perceived by the viewer. In particular, if one image is offset relative to another, there are many differences between individual pixels of one image and the corresponding pixels of the other, but these differences are not perceptually relevant provided that the boundaries are in the same relative positions. By referring to the principal boundaries of the image, rather than an absolute (pixel co-ordinate) frame of reference, any such offset can be compensated for.

The analysis may also include identification of perceptually significant image features, again identified by the shapes of the boundaries identified rather than how these boundaries are defined. The output indicative of the perceptual difference between the first and second signals can be weighted according to the perceptual significance of such image features. Significant features would include the various characteristics which make up a human face, in particular those which are significant in providing visual speech cues. Such features are of particular significance to the

human cognitive system and so errors such as distortion, absence, presence of spurious elements or changes in relative position are of greater perceptual relevance in those features than in others.

In an image containing text, those features which distinguish one character of a typeface from another (for example the serif on a letter "G" which distinguishes it from a "C") are perceptually significant.

An embodiment of the invention will now be described, by way of example only, with reference to the Figures, in which:

Figure 1 illustrates schematically a first, sensory emulation, stage of the system

Figure 2 illustrates the filter parameters used in the sensory emulation stage

Figure 3 illustrates schematically a second, perceptual, stage of the system

Figures 4, 5, 6 and 7 illustrate four ways in which boundaries may be perceived.

In this embodiment the measurement process comprises two stages, illustrated in Figures 1 and 3 respectively. The first – the sensory emulation stage – accounts for the physical sensitivity of the human visual system to given stimuli. The second – the perceptual stage – estimates the subjective intrusion caused by the remaining visible errors. The various functional elements shown in Figures 1 and 3 may be embodied as software running on a general-purpose computer.

The sensory stage (Figure 1) reproduces the gross psychophysics of the sensory mechanisms:

(i) spatio-temporal sensitivity known as the human visual filter, and

(ii) masking due to spatial frequency, orientation and temporal frequency.

Figure 1 gives a representation of the sensory stage, which emulates the physical properties of the human visual system. The same processes are applied to both the original signal and the degraded signal: these may be carried out simultaneously in parallel processing units, or they may be performed for each signal in turn, using the same processing units.

The sensory stage identifies whether details are physically perceptible, and identifies the degree to which the visual system is sensitive to them. To do so, it emulates the two main characteristics of the visual system that have an influence on the physical perceptibility of a visual stimulus:

- sensitivity of the eye/brain system
- masking effects – that is the variation in perceptual importance of one stimulus according to the presence of other stimuli.

Each of these characteristics has both a time and a space dimension, as will now be discussed.

Each signal is first filtered in temporal and spatial frequency by a filter 12, to produce a filtered sequence. The values used in the filter 12 are selected to emulate the human visual response, as already discussed in relation to Figure 2. This filter allows details that are not visible to a human visual (eye/brain) system to be removed, and therefore not counted as errors, while the perceptibility of details at other spatial and temporal frequencies is increased by the greater sensitivity of the human sensory system at those frequencies. This has the effect of weighting the information that the signals contain according to visual acuity.

The human visual system is more sensitive to some spatial and temporal frequencies than others. Everyday experience teaches us that we cannot see details smaller than a certain size. Spatial resolution is referred to in terms of spatial frequency, which is defined by counting the number of cycles of a sinusoidal pattern present per degree subtended at the eye. Closely spaced lines (fine details) correspond to high spatial frequencies, while large patterns correspond to low spatial frequencies. Once this concept is introduced, human vision can be compared to a filter, with peak (mid-range) sensitivity to spatial frequencies of around 8 cycles/degree and insensitivity to high frequencies (more than 60 cycles/degree). A similar filter characteristic can be applied in the temporal domain, where the eye fails to perceive flickering faster than about 50 Hz. The overall filter characteristic for both spatial and temporal frequency can be represented as a surface, as shown in Figure 2, in which the axes are spatial and temporal frequency (measured in cycles/degree and Hertz respectively). The vertical axis is sensitivity, with units normalised such that maximum sensitivity is equal to 1.

The second aspect of vision to be modelled by the sensory stage is known as "masking", the reduced perceptibility of errors in areas of an image where there is greater spatial activity present, and the temporal counterpart of this effect decreases the visibility of details as the rate of movement increases. Masking can be understood by considering the organisation of the primary cortex, the first stage of the brain

responsible for visual processing. Each part of the cortex is sensitive to a certain region of the retina. The incoming image stream is divided into groupings (known as channels) of spatial frequency, temporal frequency and orientation. The "next stage" of the brain processes the image stream as a set of channels, each accounting for a combination of spatial/temporal frequency and orientation in the corresponding area of the retina. Once a given channel is excited, it tends to inhibit its neighbours, making it more difficult to detect other details that are close in proximity, spatial or temporal frequency, or orientation.

Masking is a measure of the amount of inhibition a channel causes to its neighbours. This information is obtained by studying the masking produced by representative samples of channels, in terms of spatial/temporal frequency, and orientation characteristics. For the sensory stage to simulate activity masking, it is necessary to know the amount of activity present in each combination of spatial frequency and orientation within an image. This calculation can be performed using a Gabor function, a flexible form of bandpass filter, to generate respective outputs 14 in which the content of each signal is split by spatial frequency and orientation. Typically, sixteen output channels are used for each output signal, comprising four spatial orientations (vertical, horizontal, and the two diagonals) and four spatial frequencies. The resulting channels are analysed by a masking calculator 15. This calculator modifies each channel in accordance with the masking effect of the other channels; for example the perceptual importance of a low spatial-frequency phenomenon is reduced if a higher frequency spatial phenomenon is also present. Masking also occurs in the temporal sense – certain features are less noticeable to the human observer if other effects occur within a short time of them.

Calibration of this model of masking requires data describing how spatial/temporal frequency of a given orientation decreases the visibility of another stimulus. This information cannot be obtained as a complete description as the number of combinations is very large. Therefore, the separate influence of each parameter is measured. First the masking effect of a background on a stimulus is measured according to the relative orientation between the two. Then the effect of spatial and temporal frequency difference between masker and stimulus is measured. Finally, the two characteristics are combined by interpolating between common measured points.

In a simple comparison between original and degraded frames, certain types of error, such as a horizontal/vertical shift, result in large amounts of error all over the frame, but would not be noticeable to a user. This problem can be addressed by employing frame realignment, as specified in the ITU-T *"Draft new recommendation on multimedia communication delay, synchronisation, and frame rate measurement", COM 12-29-E, December 1997.* However this simple method does not fully account for the subjectivity of the error, since it does not allow for other common defects such as degradation of elements in the compressed sequence.

Following the sensory stage, the image is decomposed to allow calculation of error subjectivity by the perceptual stage (Figure 3), according to the importance of errors in relation to structures within the image. If the visible error coincides with a critical feature of the image, such as an edge, then it is more subjectively disturbing. The basic image elements, which allow a human observer to perceive the image content, can be thought of as a set of abstracted boundaries. These boundaries can be formed by colour and luminance differences, texture changes and movement as well as edges, and are identified in the decomposed image. Even some "Gestalt" effects, which cause a boundary to be perceived where none actually exists, can be algorithmically measured to allow appropriate weighting.

These boundaries are required in order to perceive image content and this is why visible errors that degrade these boundaries, for example by blurring or changing their shape, have greater subjective significance than those which do not. The output from the perceptual stage is a set of context-sensitive error descriptors that can be weighted differently to map to a variety of opinion criteria.

In some instances, a boundary may be completely absent, or a spurious boundary may be present, for example when a "ghost" image is formed by multipath reflection. In this case, the presence or absence of the boundary itself is the error.

Figure 3 is a representation of the perceptual stage, which measures the subjective significance of any errors present in the image sequence. The original signal 16 and the degraded signal 16d, each filtered and masked as described with reference to Figure 1, are first each analysed (either in parallel or sequentially) in a component extraction process 31 to identify characteristics of the edges or boundaries of the principal components of each image. These characteristics are supplied as inputs 32, 32d to a comparison process 33 which generates an output 38

indicative of the overall perceptual degradation of the degraded image with respect to the original image.

The components identified by the extraction process 31 may be distinguished by:

- Luminance (illustrated in Figure 4) and Colour
- Strong Edges (illustrated in Figure 5)
- Closure Effects (illustrated in Figure 6)
- Texture (illustrated in Figure 7)
- Movement
- Binocular (Stereoscopic) Disparities.

These last two effects rely on phenomena relating to movement and stereoscopy, not readily illustrated on the printed page. For similar reasons, only luminance differences, and not colour differences, are illustrated in Figure 4.

Figures 4 to 7 all depict a circle and a square, the square obscuring part of the circle. In each case, the boundary between the two elements is readily perceived, although the two elements are represented in different ways. In Figure 4, the circle and square have different luminance – the circle is black and the square is white. A boundary is perceived at the locations where this property changes. It will be noted that in Figures 5, 6 and 7 there are also locations where the luminance changes, (for example the boundaries between each individual stripe in Figure 7) but these are not perceived as the principal boundaries of the image.

Figure 5 illustrates a boundary defined by an edge. A "strong edge", or outline, is a narrow linear feature, of a colour or luminance contrasting with the regions on either side of it. The viewer perceives this linear feature not primarily as a component in its own right, but as a boundary separating the components either side of it. In analysis of the image, such an edge can be identified by a localised high-frequency element in the filtered signal. Suitable processes identifying edges have been developed, for example the edge extraction process described by S M Smith and J M Brady in *"SUSAN – A new approach to low-level image processing" (Technical Report TR95SMS1c, Oxford Centre for Functional magnetic Resonance Imaging of the Brain, 1995).*

In many circumstances a viewer can perceive an edge where no continuous line is present. An example is shown in Figure 6, where the lines are discontinuous.

The human perceptual system carries out a process known as "closure", which tends to complete such partial edges. (A further example is illustrated by the fact that none of Figures 4 to 7 actually depict a full circle. The viewer infers the presence of a circle from the four lenticular regions actually depicted in each Figure). Various processes have been developed to emulate the closure process carried out by the human perceptual system. One such process is described by Kass M., Witkin A., and Terzopoulos D., *"Snakes: Active Boundary Models"*, published in the *Proceedings of First International Conference on Computer Vision 1987*, pages 259-269.

"Texture" can be identified in many regions in which the properties already mentioned are not constant. For example, in a region occupied by parallel lines, of a colour or luminance contrasting with the background, the individual location of each line is not of great perceptual significance. However, if the lines have different orientations in different parts of the region, an observer will perceive a boundary where the orientation changes. This property is found for instance in the orientation of brushstrokes in paintings. An example is shown in Figure 7, in which the circle and square are defined by two orthogonal series of parallel bars. Note that if the image is enlarged such that the angular separation of the stripes is closer to the peak value shown in Figure 2, and the dimensions of the square and circle further from that peak value, the individual stripes would become the dominant features, instead of the square and circle. It will also be apparent that if the orientations of the bars were different, the boundary between the square and the circle may become less distinct. To identify the texture content of a region of the image, the energy content in each channel output from the Gabor filters 13 is used. Each channel represents a given spatial frequency and orientation. By identifying regions where a given channel or channels have high energy content, regions of similar texture can be identified.

Shapes can be discerned by the human perceptual system in other ways, not illustrated in the accompanying drawings. In particular, disparities between related images, such as the pairs of image frames used in stereoscopy, or successive image frames in a motion picture, may identify image elements not apparent on inspection of a single frame. For example, if two otherwise similar images, with no discernible structure in either individual image, include a region displaced in one image in relation to its position in the other, the boundaries of that region can be discerned if the two images are viewed simultaneously, one by each eye. Similarly, if a region of

apparently random pixels moves coherently across another such region in a moving image, that region will be discernible to an observer, even though no shape would be discernible in an individual frame taken from the sequence. This phenomenon is observable in the natural world – there are many creatures such as flatfish which have colouring similar to their environment, and which are only noticeable when they move.

The component extraction process identifies the boundaries of the principal elements of both the original and degraded signals. The perceptual importance of each boundary depends on a number of factors, such as its nature (edge, colour change, texture, etc), the degree of contrast involved, and its context. In this latter category, a high frequency component to the filtered and masked signal will signify that there are a large number of individual edges present in that region of the image. This will reduce the significance of each individual edge – compare Figure 5, which has few such edges, with Figure 7, which has many more such edges.

Each individual extraction process carried out in the component splitting step 31, on its own, typically performs relatively poorly, as they all tend to create false boundaries, and fail to detect others. However, the combination of different processes increases the quality of the result, a visual object being often defined by many perceptual boundaries, as discussed by Scassellati B.M. in *"High-level perceptual contours from a variety of low-level physical features"* (Master Thesis, Massachusetts Institute of Technology, May 1995). For this reason the comparison process 33 compares all the boundaries together, regardless of how they are defined except insofar as this affects their perceptual significance, to produce a single aggregated output 38.

The results 32, 32d of the component analysis 31 are passed to a comparison process 33, in which the component boundaries identified in each signal are compared. By comparing the perceptual relevance of all boundary types in the image, a measure of the overall perceptual significance of degradation of a signal can be determined, and provided as an output 38. The perceptual significance of errors in a degraded signal depends on the context in which they occur. For example, the loss or gain of a diagonal line (edge) in Figure 7 would have little effect on the viewer's perception of the image, but the same error, if applied to Figure 5, would have a

much greater significance. Similarly, random dark specks would have a much greater effect on the legibility of Figure 6 than they would on Figure 4.

In more detail, the comparison process 33 consists of a number of individual elements. The first element identifies the closest match between the arrangements of the boundaries in the two images (34), and uses this to effect a bulk translation of one image with respect to the other (35) so that these boundaries correspond.

The next process 36 identifies features to which the human cognitive system is most sensitive, and weighting factors W are generated for such features. For example, it is possible to weight the cognitive relevance of perceptually critical image elements such as those responsible for visual speech cues, as it is known that certain facial features are principally responsible for visual speech cues. See for example Rosenblum, L.D., & Saldaña, H.M. (1996). *"An audiovisual test of kinematic primitives for visual speech perception"*. (Journal of Experimental Psychology: Human Perception and Performance, vol 22, pages 318-331) and Jordan, T.R. & Thomas, S.M. (1998). *"Anatomically guided construction of point-light facial images"*. (Technical report: Human Perception and Communication Research Group, University of Nottingham, Nottingham, U.K).

We can infer that a face is present using pattern recognition or by virtue of the nature of the service delivering the image.

The perceptual significance of each boundary in one image is then compared with the corresponding boundary (if any) in the other (37), and an output 38 generated according to the degree of difference in such perceptual significance and the weightings W previously determined. It should be noted that differences in how the boundary is defined (hard edge, colour difference, etc) do not necessarily affect the perceptual significance of the boundary, so all the boundaries, however defined, are compared together. Moreover, since the presence of a spurious boundary can be as perceptually significant as the absence of a real one, it is the absolute difference in perceptibility that is determined.

Note that degradation of the signal may have caused a boundary defined by, for example, an edge, to disappear, but the boundary may still be discernible because of some other difference such as colour, luminance or texture. The error image produced by established models (filtered and masked noise) provides an indication of the visible degradation of the image. The comparison process 37 includes a measure

of the extent to which the essential content is maintained and offers an improved measure of the image intelligibility. In comparing the boundaries (step 37), the perceptual significance of a given boundary may depend on its nature. A boundary between different textures may be less well defined than one defined by an edge, and such reduced boundary perceptibility is taken into account in generating the output.

This process is suitable for great range of video quality assessment applications, where identification and comparison of the perceptual boundaries is necessary. A good example is given by very low bandwidth systems where a face is algorithmically reconstructed. This would be impossible for many of the previously known visual models to assess appropriately. The comparison of perceptual boundaries also enables the assessment of synthetic representations of images such as an animated talking face, in which the features of the image that facilitate subsequent cognitive interpretation as a face are of prime importance.

CLAIMS

1. A method of measuring the differences between a first video signal (16) and a second video signal (16d), comprising the steps of:

analysing (31) the information content of each video signal to identify the perceptually relevant boundaries of the video images depicted therein;

comparing (33) the boundaries so defined in the first signal with those in the second signal; the comparison including determination of the extent to which the properties of the boundaries defined in the original image are preserved, and

generating an output (38) indicative of the perceptual difference between the first and second signals.

2. A method according to Claim 1, in which the information content is analysed for a plurality of boundary-identifying characteristics (32, 32d), and the properties of the boundaries on which the comparison (37) is based include the characteristics by which such boundaries are defined in each of the signals.

3 A method according to claim 2, wherein the characteristics include the presence of edges.

4 A method according to claim 2 or 3, wherein the characteristics include the presence of disparities between frames

5 A method according to claim 2, 3 or 4, wherein the characteristics include changes in at least one of the properties of: luminance, colour or texture.

6 A method according to any of claims 1 to 5, in which the comparison includes a comparison (36) of the perceptibility of corresponding boundaries identified in the first and second signals.

7 A method according to any preceding claim, in which the comparison of the images includes the steps of

identification (34) of the principal elements in each image, and

compensation (35) for differences in the relative positions of the said principal elements.

8 A method according to any preceding claim, in which the analysis includes identification of perceptually significant image features, and the output (38) indicative of the perceptual difference between the first and second signals is weighted according to the perceptual significance of such image features.

9. A method according to claim 8, in which the perceptually significant image features are those characteristic of the human face.

10. A method according to claim 9, in which a weighting is applied to the output according to the significance of the feature in providing visual cues to speech.

11 A method according to claim 8, in which the perceptually significant image features are those by which individual text characters are distinguished.

12. Apparatus for measuring the differences between a first video signal (16) and a second video signal (16d), comprising:

analysis means (31) for the information content of each video signal to identify the perceptually relevant boundaries of the video images depicted therein;

comparison means (33) for comparing the boundaries so defined in the first signal (16) with those in the second signal (16d); the comparison including determination of the extent to which the properties of the boundaries defined in the original image are preserved,

and means for generating an output (38) indicative of the perceptual difference between the first and second signals (16, 16d).

13. Apparatus according to Claim 12, wherein the analysis means (31) is arranged to analyse the information content in the signals (16, 16d) for a plurality of boundary-identifying characteristics (32, 32d), and the comparison means (33) is arranged to compare the characteristics by which such boundaries are defined in each of the signals.

14 Apparatus according to claim 13, wherein the analysis means (31) includes means to identify the presence of edges.

15. Apparatus according to claim 13 or 14, wherein the analysis means (33) includes means to identify the presence of disparities between frames

16. Apparatus according to claim 13, 14 or 15, wherein the analysis means (33) includes means to identify differences in at least one of the properties of: luminance, colour or texture.

17 Apparatus according to any of claims 12 to 16, in which the comparison means (33) includes means (36) for determining the perceptibility of the boundaries identified in the first and second signals.

18 Apparatus according to any of claims 12 to 17, in which the comparison means (33) includes image matching means (34) for identification of the principal elements in each image and translation means (35) for effecting translation of one image (16d) to compensate for differences in the relative positions of such elements in the first and second images.

19. Apparatus according to any of claims 12 to 18, in which the comparison means (33) includes weighting means 36 for identifying perceptually significant image features in the components (32, 32d), and weighting the output (38) according to the perceptual significance of such image features.

20. Apparatus according to any of claims 12 to 19, further comprising visual stage means (11,12,13,14,15) for processing original input signals (11) to emulate the response of the human visual system, to generate modified input signals (16, 16d) for input to the analysis means (31).

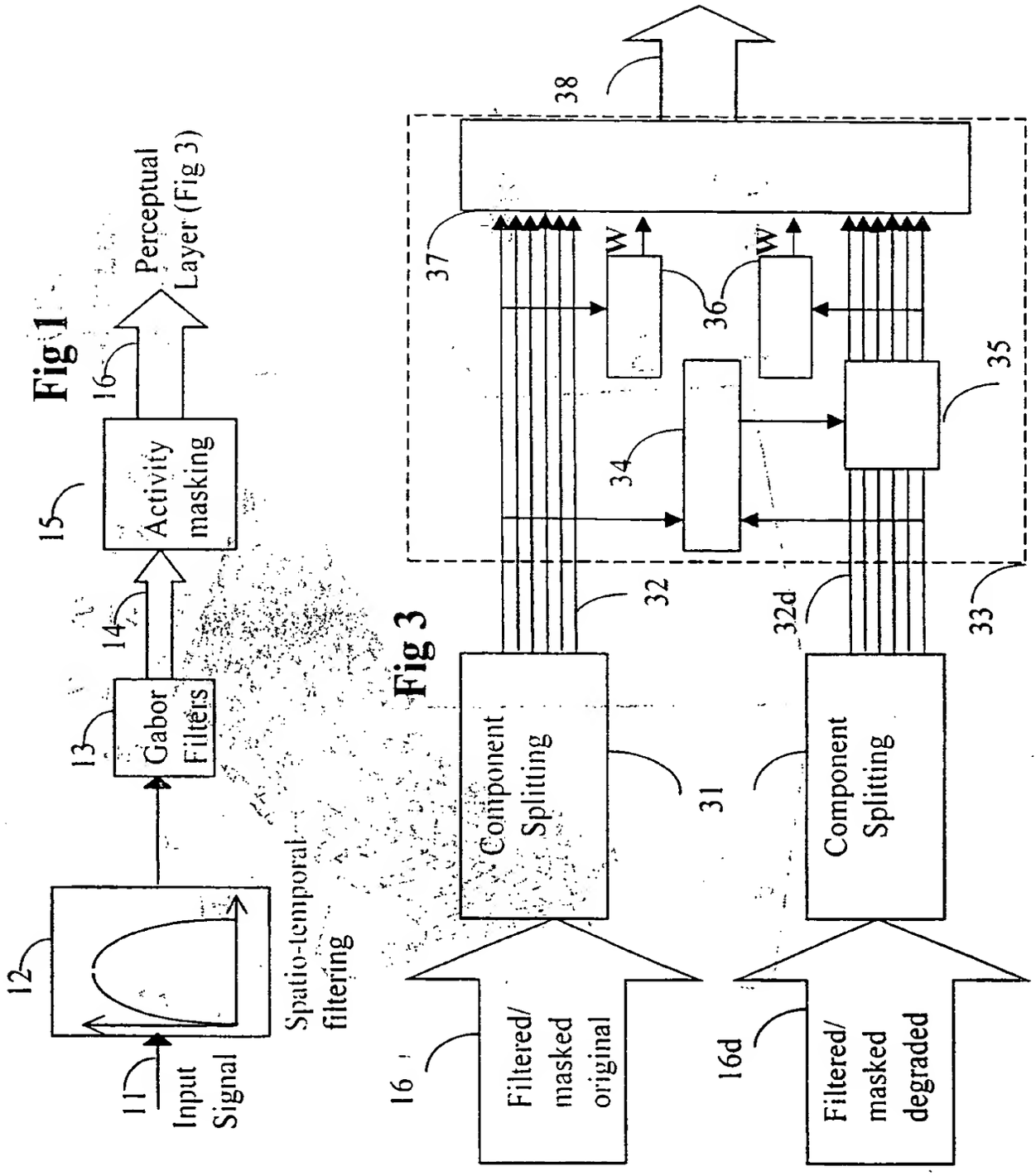
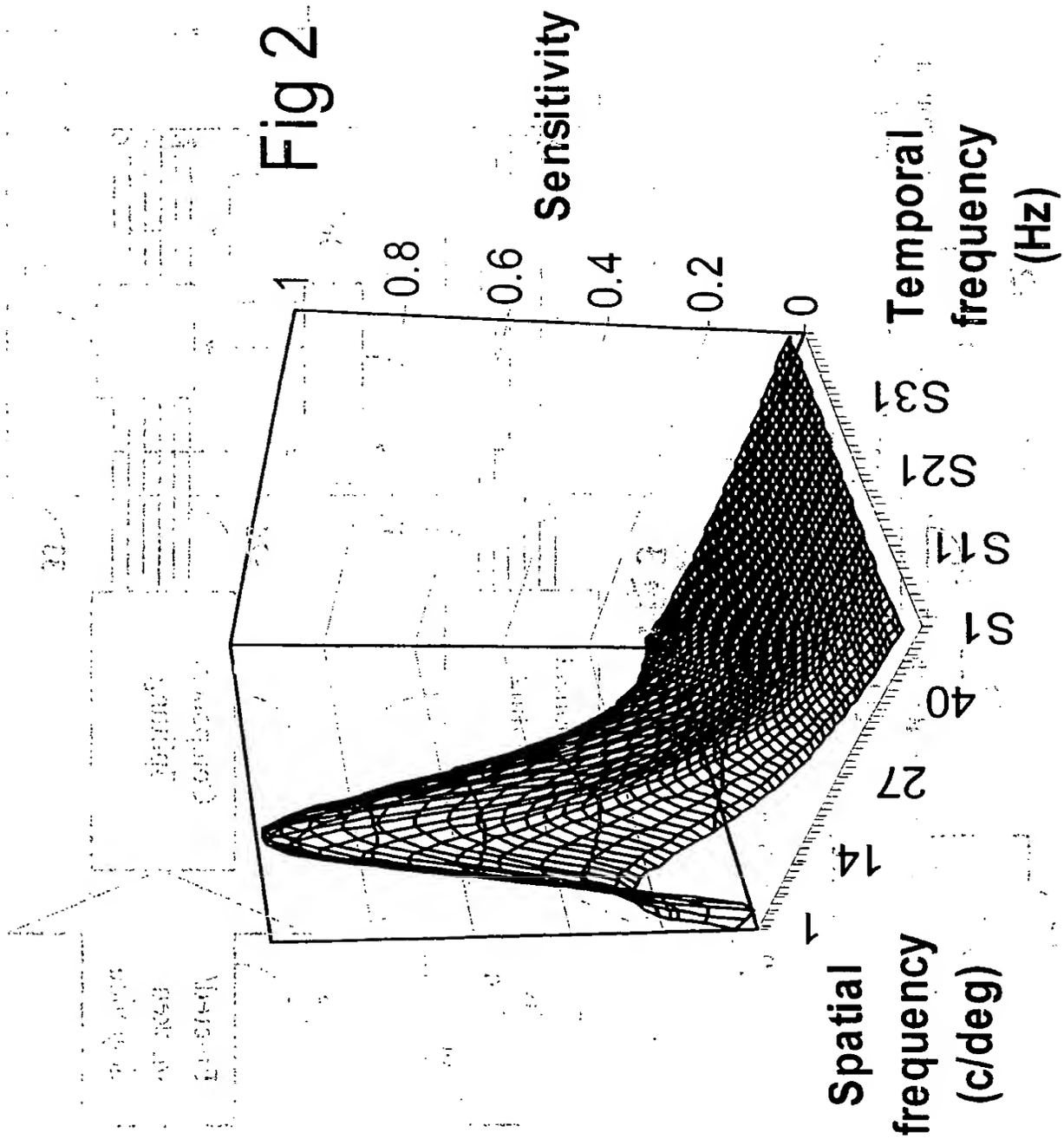


Fig 2



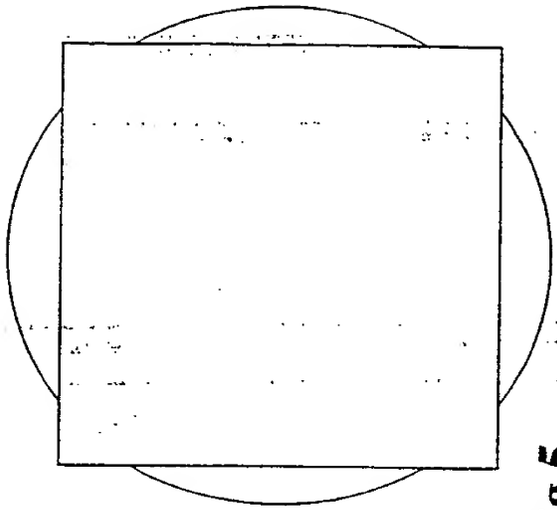


Fig 5

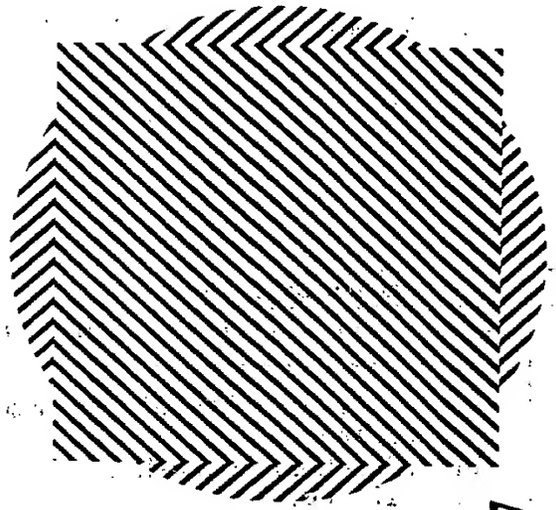


Fig 7

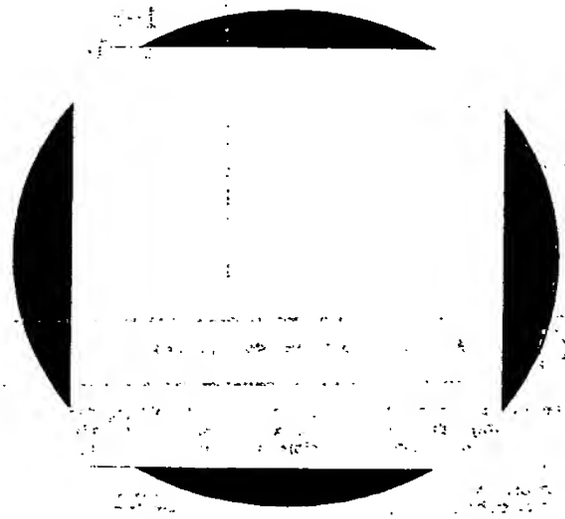


Fig 4

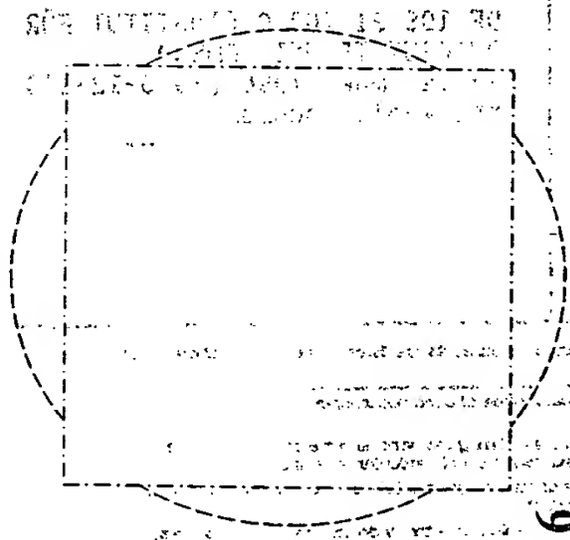


Fig 6

INTERNATIONAL SEARCH REPORT

Inte onal Application No
PCT/GB 00/00171

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 H04N17/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 H04N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,A	RIX A. ET AL: "Models of human perception" BT TECHNOLOGY JOURNAL., vol. 17, no. 1, 19 March 1999 (1999-03-19), pages 24-34, XP000824576 BT LABORATORIES., GB ISSN: 0265-0193 the whole document	1-20
A	DE 195 21 408 C (INSTITUT FÜR RUNDFUNKTECHNIK GMBH) 12 December 1996 (1996-12-12) the whole document	1-3, 12-14

-/--

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
24 February 2000	02/03/2000

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Verschelden, J

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Inte. onal Application No
PCT/GB 00/00171

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
A	TAN K. ET AL: "An objective measurement tool for MPEG video quality" SIGNAL PROCESSING. EUROPEAN JOURNAL DEVOTED TO THE METHODS AND APPLICATIONS OF SIGNAL PROCESSING., vol. 70, 1998, pages 279-294, XP004144970 ELSEVIER SCIENCE PUBLISHERS B.V. AMSTERDAM., NL ISSN: 0165-1684 the whole document	1,12
A	US 5 446 492 A (WOLF S. ET AL) 29 August 1995 (1995-08-29) the whole document	1,12

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Information on patent family members

Inte onal Application No
PCT/GB 00/00171

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
DE 19521408 C	12-12-1996	NONE	
US 5446492 A	29-08-1995	US 5596364 A	21-01-1997